DOCUMENT RESUME

ED 360 412 UD 029 214

AUTHOR	Stallings, Jane; Freiberg, H. Jerome
TITLE	Observation for the Improvement of Teaching. Publication Series 91-3.
INSTITUTION	Temple Univ., Philadelphia. Center for Research in Human Development and Education.
PUB DATE	91
NOTE	34p.; In: Waxman, Hersholt C. and Walberg, Herbert J. "Effective Teaching: Current Research." Berkeley, CA, McCutchan Publishing Corporation, 1991.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Classroom Observation Techniques; Classroom Research; Early Childhood Education; Economically Disadvantaged; *Educational Improvement; Elementary Secondary Education; English (Second Language); *Program Effectiveness; *Staff Development; *Student Evaluation; Student Teaching; Time Management; Training
IDENTIFIERS	*Project Follow Through

ABSTRACT

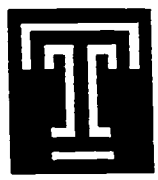
The use of classroom observation is explored in several capacities. Specific observation instruments that were developed to evaluate the effectiveness of the National Follow Through Program were later used (sometimes in adapted forms) to study early childhood programs, secondary school programs, student teacher effectiveness; and use of time across a school district. Project Follow Through was intended to provide a program analogous to Head Start for economically disadvantaged children over a longer period of time. This chapter presents the fourth and most comprehensive report of Follow Through classroom observation data collected in spring 1973 from 36 sites representing 7 sponsors and 7 program models (35 first grades and 36 third grades). Other studies using the developed Classroom Observation Instrument (COI) in the following are briefly outlined: (1) early childhood education; (2) English as a Second Language; (3) secondary school; (4) staff development; (5) effective use of time training; (6) student evaluation; and (7) student teaching. The observation techniques have provided a means to identify effective instructional practices in a wide range of classroom settings. Two tables and one figure illustrate the discussion. (SLD)

Temple University Center for Research in Human Development and Education

Publication Series 91-3

Observation for the Improvement of Teaching

Jane Stallings and H. Jerome Freiberg



Observation for the Improvement of Teaching

Jane Stallings
Professor, College of Education
Texas A&M University

H. Jerome Freiberg

Professor of Curriculum and Instruction
University of Houston
Research Associate
Temple University Center for Research
in Human Development and Education

Abstract

This chapter explores the use of classroom observations in several capacities. Specific observation instruments that were developed to evaluate the effectiveness of the National Follow Through program were later used (sometimes in adapted forms) to study early childhood programs, secondary school programs, student teacher effectiveness, and use of time across a school district. A study was also conducted to evaluate the effects of using observation data as a basis for staff development.

5

Observation for the Improvement of Teaching

Jane Stallings and H. Jerome Freiberg

INTRODUCTION

In 1969, Stanford Research Institute (SRI) held a contract with the National Institute of Education to evaluate National Head Start and Follow Through Planned Variation programs. Project Follow Through was established by the U. S. Congress in 1967 (the legislative authority was the Economic Opportunity Act of 1964, as amended) when it became apparent that a program was needed in the early grades of public school that was articulated with Project Head Start . als and approaches and, therefore, would provide a comparable educational program for economically disadvantaged children over a longer period of time. A clearly stated purpose of the Follow Through program was the enhancement of the life chances of the economically deprived child.

According to Deutsch (1967), "children from backgrounds of social marginality enter the first grade already behind their middle-class counterparts in a number of skills highly related to scholastic achievement. They are simply less prepared to meet the demands of the school and the classroom situation. . . . In other words, intellectual and achievement differences between lower-class and middle-

class children are smallest at the first-grade level, and tend to increase through the elementary school years" (pp. 64–65). However, an evaluation by Wolff and Stein (1966) of the first summer program of Head Start in 1965 had indicated that the initial achievement gains of the children had not been maintained in the public school. These early findings were believed to indicate that a more sustained program of longer duration might produce lasting gains. The result was the establishment of Follow Through Planned Variation as a longitudinal quasi-experimental program that would evaluate the ability of an intervention program to enhance the educational achievement of economically disadvantaged children.

Project Follow Through was originally set up in a "planned variation" research design; that is, the goal was to examine the differential effectiveness of programs based on divergent educational and developmental theories. The program began when researchers and other educational stakeholders were invited by the government to submit plans for establishing their various programs in public schools in order to test whether their individual approaches could improve the educational achievement of economically disadvantaged children. From the group that came forward, twenty-two were selected to implement their programs as Follow Through sponsors. We refer to "sponsors" here as those responsible for constructing and implementing the educational programs (or models). Eleven of the twenty-two sponsors of educational programs had developed and tried their models in university settings, eight were affiliated with private research institutes, and three were community-developed programs.

The sponsors of educational programs described their models to an audience that included representatives from school districts around the country at a conference in Kansas City, Kansas, in 1968. Ultimately, these models were implemented in 154 Follow Through projects within 136 urban and rural communities throughout the nation. The Follow Through sponsors then faced the challenge of program implementation, including guiding the behavior of teachers toward specified goals set by the sponsors. Egbert (1973) provides a historical view of the Follow Through project.

In other evaluations of Follow Through Planned Variation, the major emphasis was to determine if the models affected children's performance. Yet it was clear that if such effects were found, and if the effects were different from one model to another, we would not know what caused the differences. Therefore, we needed to know what was

actually happening in the classrooms. In order to determine whether the sponsors were effective in getting teachers to practice their methods in the classroom, it was necessary to observe the classrooms systematically. We wanted to know whether a child's day in the classroom corresponded with the sponsor's educational prescriptions. To assess this, we needed a comprehensive observation instrument.

In the fall of 1969, the SRI staff, with assistance from twelve sponsors' representatives, developed an observation system with which a wide range of classroom behaviors could be recorded and with which objective information could probably be recorded that would provide a fair assessment of all sponsors' models. The procedures that were developed could record activities, materials used, groupings, and interactions. This chapter presents the fourth and most comprehensive report of Follow Through classroom observation data. The data for this study were collected in spring 1973 from thirty-six sites representing seven sponsors. The seven models of the chosen sponsors represent a wide spectrum of innovative educational theories and were selected because each model was being implemented in at least five locations. The models selected for this study include two models based on positive reinforcement theory (from the University of Kansas and from the University of Oregon), a model based primarily on the cognitive developmental theory of Jean Piaget (High/Scope Foundation), an open-classroom model based on the English Infant School Theory (Education Development Center), and three other models drawn from Piaget, John Dewey, and the English Infant Schools (Far West Laboratory, the University of Arizona, and Bank Street College).

The study focuses on whether sponsors can deliver their educational models to diverse communities, and explores the effects of training classroom personnel to use specific procedures in the classroom. An educational theory can be proved effective only if the teachers and aides carry out program specifications. Such specifications set by sponsors include the physical arrangement of the classroom, utilization of the prescribed curricula, and interactions with children. This study addresses the following issues:

1. Are the observed classroom programs consistent with their sponsors' stated intentions? That is, does the model show a relatively high frequency of occurrence of those elements of the program that the sponsor rated as important?

2. Are the sponsored classrooms consistent within a site and between sites? That is, do the four third-grade classrooms at site A score similarly on specific program components and do they resemble the third-grade classrooms at sites B, C, D, and E?
3. How do selected classroom processes relate to scores on the following: Metropolitan Achievement Test (MAT) (reading and arithmetic), Raven's Coloured Progressive Matrices, Coopersmith's Self-Esteem Inventory, and Intellectual Achievement Responsibility Scale (IAR)?

SAMPLE

Four first-grade and four third-grade classrooms were observed in each of thirty-six cities and towns. These represented five projects for each of six of the Follow Through educational models and six projects for the University of Arizona's model. One first-grade and one third-grade non-Follow Through classroom were selected for comparison at each project site. These non-Follow Through classrooms were combined to form two pooled comparison groups, thirty-five first grades and thirty-six third grades. The projects included in the sample represented all geographic regions, urban and rural areas, and several racial and ethnic groups.

Observation sites were selected according to the following criteria: (1) they were among the sites where pupil testing was to occur in spring 1973 as part of the Follow Through evaluation; (2) each sponsor would, as much as possible, have a balanced geographic distribution of sites, which included urban-rural and north-south projects; and (3) each sponsor would have included at least two sites where he thought the model was well implemented.

In addition to identifying classrooms for observation, we randomly selected four children from each classroom for individual observations. At each site, the primary consideration in identifying the classrooms and children to be observed was the availability of baseline data for the children when they entered school in kindergarten or first grade.

In those projects where baseline data were not available, the Follow Through classrooms were nominated by the sponsor, and the non-Follow Through classrooms were selected by the SRI Field

Operations staff. The SRI staff selected children for individual observation on a random basis from classroom roster lists.

MEASUREMENT INSTRUMENTS

Behavior Observation

The Classroom Observation Instrument (COI) is designed to record classroom arrangements and elements of events considered educationally significant by the Follow Through sponsors.

Formation of Variables. Many of the individual codes are too molecular to serve effectively as measures of classroom educational characteristics. Hence, it was necessary to form theoretically significant variables by combining certain codes. The COI consists of 602 categories describing behaviors of teachers and children in the classroom situation. The items identify materials used in the classroom, the grouping arrangements of teacher and children, the activities that occur, the behavior of an individual, and the interactions that occur between two or more people.

Interaction observations were made in five-minute sequences. A form of shorthand was used to record the continuous action and interaction of selected persons in the classrooms. On two of three days of observation, there was an adult focus, that is, the classroom adults were the subjects of observation. On the remaining day, the four randomly selected children were the focus. Hence, the data provide one set of measures of classroom process (adult focus) and one set of child behaviors (child focus) with the same set of categories, or codes.

Observers were instructed to complete approximately four observations each hour during the five-hour observation day; hence, it was hoped that a total of twenty observations would be completed each day, or sixty for each classroom. For the 1,011 observation days of all observers, the adjusted mean number of observations completed each day was 18.88, with a standard deviation of 2.17. Fewer observations occurred for certain classrooms because of intervening events during the class day that prohibited observations. Data from any day that had fewer than twelve observations were deleted.

The data were collected on three consecutive days in spring 1973. In most cases, the teachers had been working with the sponsors'

educational models for two or three years. No beginning teachers were selected for observation.

Since there are over 100,000 possible combinations of codes that could form variables, it was important to formulate only those variables considered relevant to the study of sponsor implementation. The following sections will describe the transformation of codes from each portion of the classroom observation instrument into variables.

Because this study evaluated classroom environments and classroom instructional processes, the classroom rather than the child was the unit of analysis. Classroom mean scores were also computed for the sample of individually observed children. Each classroom was assigned a value on a given variable based on the sum of the frequency of occurrence of the variable for the observation days.

Classroom Summary Information (CSI) Variables. Once a day, before the observation with the COI started, the observer recorded information that identified the classroom by sponsor, site, teacher, grade, and observer. The observer also noted the numbers of adults and children present. To obtain the ratio of children to teachers and aides, the total number of children present on each observation day was divided by the total number of teachers and aides present. An average ratio over the three days was then computed. Total class duration was computed by averaging the number of class hours recorded for the three days of observations.

Physical Environment Information (PEI) Variables. This section of the COI, completed once each observation day, provided two kinds of information: (1) seating and workgroup patterns, and (2) equipment and materials present and used in the classroom. The scores for a classroom were based on the sum of all three days.

Classroom Checklist (CCL) Variables. The CCL variables define the frequency of occurrence of specific activities (e.g., group time, mathematics, dramatic play) that denote the frequency of occurrence of the different groupings of adults and children (e.g., aide with small group of children, one child without an adult), grouping within particular activities (e.g., teacher with two children in mathematics activity), and the use of special materials or equipment (e.g., texts or workbooks, audiovisual equipment) within the activities of mathematics, reading, s udies, and science. Some metavariables were formed,

such as "How frequently does a child receive individual attention from an adult?" This metavariable was formed by adding many discrete subvariables, such as "How frequently does a child receive individual attention from an aide during mathematics?" plus "How frequently does a child receive individual attention from a volunteer during reading?" plus all other variables that describe an incident where one child is working with an adult.

Five-minute Observation (FMO) Variables. This main portion of the COI is used to record, in the form of coded sentences, interactions that occur in the classroom. The Flanders Interaction Analysis Observation system served as the model for this section of the COI. The Who, to Whom, What, and How codes have functions and operational definitions similar to the Flanders system (1970). For this purpose, the observer used a series of four-celled frames (see Figure 5-1 for frames used in preschool or kindergarten classrooms). To record each interaction, the observer made a check mark in the appropriate circle in each of the four cells of a frame. These marks identified the speaker, the person being spoken to, and the message

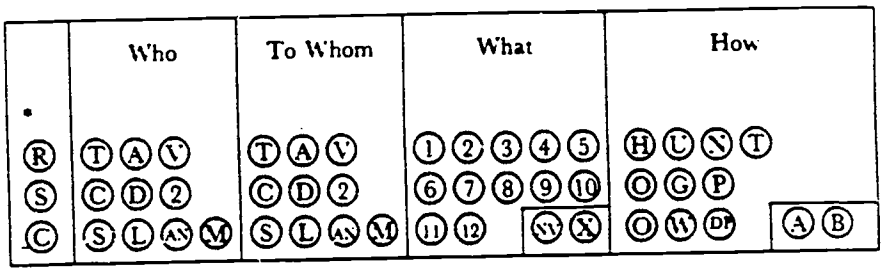


Figure 5-1
Frames Used for Observations in Preschool and Kindergarten Classrooms

Who and *To Whom* categories: T–teacher, A–aide, V–volunteer, C–child. D–different child, 2–two children, S–small group, L–large group, An–animal, M–machine. *What* categories: 1–command or request; 1Q–direct question (Q from *How* col.); 2–open-ended question; 3–response; 4–instruction, explanation; 5–general comments/general action; 6–task-related comment; 7–acknowledge; 8–praise; 9–corrective feedback; 10–no response; 11–waiting; 12–observing, listening; NV–nonverbal; X–movement. *How* categories: H–happy; U–unhappy; N–negative; T–touch, Q–question; G–guide/reason; P–punish; O–object; W–worth; DP–dramatic play/pretend; A–academic; B–behavior.

• R–Repeat the Frame, S–Using Second Language, C–Cancel the Frame

being delivered. The How column describes the emotional affect and whether the conversation had an academic content or referred to behavior. Each frame represents a sentence. If one person asks a question, it is coded in the first frame. A second frame is used for the response of the other person, and so on.

For example, the following three interactions would require three frames:

1. TEACHER: Maria, what did you like best about the story *Peter Pan*?
   [In our shorthand, this sentence is coded TC2QA. The teacher (T, Who column) has asked Maria (C, To Whom column) a thought-provoking question (2Q, What column). The question is about the academic subject in the How column (A, How column).]
2. MARIA: Tinkerbell. She was very brave.
   [Shorthand: CT3A. The child (C, Who column) responds (3, What column) with academic content (A, How column) to the teacher's (T, To Whom column) question.]
3. TEACHER: Oh yes, she was brave, wasn't she?
   [Shorthand: TC7A. The teacher (T, Who column) acknowledges (7, What column) the child's (C, To Whom column) academic (A, How column) response.]

Seventy-two of these frames represent a five-minute interaction period. The variables formed from these complex codings were those that seemed most appropriate to the sponsors' models and to the analysis planned for this study.

The FMO variables were selected and named to describe interactions relevant to sponsors' implementation. The variables are defined by appropriate code combinations or sentences. Generally, the FMO variables describe child-adult verbal interactions (i.e., questions, responses, instruction, comments, and feedback) and nonverbal interactions (i.e., nonverbal requests, responses, self-instruction, feedback, waiting, and observing/listening). In some cases, these FMO variables are further defined by the How category modifiers (such as academic, social behavior, happy, negative). A few variables are defined by the sequential ordering of certain interaction frames (e.g., adult question followed by child response followed by adult feedback).

Observer Reliability

All observers were trained in a seven-day intensive training course delivered at four national training sites. Potential observers watched a specimen videotape, and only those who met the final criterion of coding interactions with a reliability of 70 percent or above were employed to collect data for this study. Of the original seventy-two who were trained, sixty-three met this criterion. Nine more observers were trained in a special session to fill these vacancies.

The observers began work in classrooms; after approximately two weeks, twenty simulated classroom situations were videotaped and shown to the observers to code. Each simulation was approximately twenty interaction frames long. These simulations attempted to present several concise, clear examples of each code used in the COI. Each simulation began with a still frame in which the narrator identified the focus of the observation (a teacher or a child). Each skit was shown with a two-second pause between interactions. The observers were instructed to code one interaction frame during each pause.

Matrices were constructed for each observer. A form was prepared that listed all sixteen What codes across the top of the matrix and down the side. Those across the top were the "true" codes as judged by the investigators; the numbers of instances of each occurring in the twenty situations were listed in the row under the labels. The codes listed down the side were the actual codes ascribed by the observer being tested. The reliability booklets for each observer were examined frame by frame, and tallies were made of each observer's coded interaction sequences. If the observer's coding agreed with the criterion, a tally was placed in the intersection of the row and column. The principal diagonal, then, contains the cells indicating the observer's correct coding; other cells contain incorrect coding. The row totals are the total number of times an observer recorded each code, whether correctly or not. The number of criterion examples shown across the top could be compared with the diagonal to compute the observer's reliability for each code. An examination of a particular cell in the row reveals whether the code was recorded correctly or incorrectly, and, if recorded incorrectly, the row cells show exactly which codes were confused with one another. (This is reported in great detail by Stallings and Gieson, 1977.)

Thus, observer bias can be assessed by examining the overuse, underuse, or confusion of codes. In this study, each observer was responsible for observing one grade level at a single location and therefore the data collected by each observer is identifiable in analysis. The value of this method for measuring accuracy is that it contributes directly toward interpreting the data.

Other Child Measures

The children's ability when they began school was assessed by the Wide Range Achievement Test (WRAT). It was administered to children when they entered school, either at the kindergarten or first-grade level.

Reading and mathematics skills were assessed by the MAT in both first and third grades. Problem-solving skills (perceptual) were assessed in third grade only, using the Raven's Coloured Progressive Matrices. This test, designed as a culture-fair, fluid intelligence test, was adopted for use in the evaluation as a measure of nonverbal reasoning and problem-solving ability in visual perceptual tasks. The Intellectual Achievement Responsibility Scale (IAR), used in third grade only, assessed the extent to which the child takes responsibility for his own successes or failures (i.e., internal locus of control) or attributes his achievements to the operation of external forces (e.g., luck or fate). Child behaviors were assessed through systematic observations recorded on the COI. Absences from school were determined from school records.

CONSISTENCY OF CLASSROOM PROCESSES

We examine in this section the day-to-day variability of what occurs in the classroom. For the purposes of this chapter, we would like to have consistent descriptions of what was occurring in a classroom during spring 1973. The activities and interactions that occur in a classroom no doubt change radically over the course of a school year as adults and pupils become acquainted, as the subject matter changes, and as holiday seasons pass. Even when inferences are confined to the spring of the school year, after the Follow Through teacher has had approximately six months to implement a sponsor's

model, classroom processes no doubt vary from day to day. It was, therefore, important to find out how stable our descriptions of these processes could be when based on only a few days of observation.

Three consecutive days of observation were scheduled for each classroom, both Follow Through and non-Follow Through. The values of the Classroom Check List (CCL) variables are based on all three days of observation; the values of the adult/activity focus Five-Minute Observation (FMO) interaction variables were based on two days of observation, while the child-focus observations were based on a single day per class.

A subset of CCL and FMO variables was chosen for the assessment of the stability of the classroom processes. The variables were selected on the basis of how well they described sponsors' programs. Results from previous evaluations (Stallings, 1973; Stallings, Baker, and Steinmetz, 1972) were used in the selection.

For each variable, the correlations were computed between the observed values on the two days of adult/activity focus observations. The Spearman-Brown formula was applied to the correlations to derive the consistency of two or three days of observation for the FMO and CCL variables, respectively. (Since there was only a single day of child-focus observations per class, the child-focus FMO variables were not included in this analysis.)

The consistency coefficient reflects the variability of the obtained classroom means, part of which is a product of "true" variance in classroom variables, while the remainder stems from measurement errors. Because of the method of determining observer reliability (measurement error), there is no satisfactory way to untangle the two by a correction for attenuation.

A primary factor contributing to less-than-perfect consistency is the assumed variability of the classroom processes from day to day. Another factor is the variability of the children's absences from day to day and differences in the number of absences across classrooms. A high-consistency coefficient, say above 0.70, indicates that the classrooms maintain approximately the same rank order on observed scores from day to day. This would indicate that error due either to day-to-day variability within classrooms or to absences is slight, although it would not rule out the possibility of systematic error operating across absences.

For all classrooms combined, both sponsored and non-Follow Through, the coefficients are reasonably high. Those for the CCL

variables are above 0.70, with the exception of variable 66 (numbers, mathematics, arithmetic) for the third grade, where the coefficient was 0.68. For the adult/activity focus FMO variables, the coefficients were all above 0.85, with the exception of variable 374a (adult instruction, academic) for first grade, where the coefficient was 0.74.

For the individual sponsors, approximately 84 percent of the 140 coefficients had a value of 0.70 or more. The reliability coefficient for variable 66 (numbers, mathematics, arithmetic) was below 0.70 in six out of the fourteen cases. In particular, the coefficients were extremely low for both grade levels of the University of Arizona and for the third grades of Bank Street and of the University of Oregon. The negative coefficient for Bank Street's third grade is the result of one classroom in which an extremely high proportion of the class time was spent in mathematics on the first day and a small proportion on the second day. The extremely low-consistency coefficients for the University of Oregon on variable 66 in the third grade and variable 67 in the first grade are notable because this sponsor's program is considered more structured than others.

In summary, the coefficients computed over all classrooms indicate that the consistency of instructional processes was surprisingly high. The differences among classrooms account for a substantial portion of the variability among the variables we have selected. The same conclusion holds with a few exceptions for the coefficients computed for each sponsor and grade level. The only variable for which the day-to-day consistency was low for several sponsors was variable 66 (average amount of time that a child was observed to be engaged in numbers, mathematics, arithmetic).

MEASUREMENTS OF APPROXIMATION TO THEORETICAL MODEL

The first step in the assessment of classroom implementation was to describe each educational model in detail. These descriptions were prepared by our staff and reviewed by the sponsors, then revised according to the sponsors' specifications. The second step was to create variables from the codes used in the observation instrument that would describe representative elements of each sponsor's model. Each sponsor identified those variables that were (1) important to his

model and (2) expected to occur more frequently than in conventional classrooms. A list of variables was made for each of the seven models. The number of variables ranged from sixteen for the University of Oregon to twenty-eight for Far West Laboratory (see Table 5-1). The critical list of variables describes a sponsor's model only in part; the observation instrument employed in the study is not designed to capture the important subtle processes of some of the programs. For example, a goal of Far West Laboratory is to have teachers establish environments where a child can search for solutions to problems in his own way and can risk, guess, and make discoveries without serious negative psychological consequences. It was not possible for us to measure directly the extent to which such an environment had been established.

Since the Follow Through programs are intended to be innovative and to represent alternatives to the conventional classroom, a pool of non-Follow Through classrooms was used as the standard from which Follow Through classrooms were expected to differ in specified ways. The standards were established separately for first and third grades.

With observational data, the distribution of scores rarely follows a normal curve; thus a nonparametric scaling technique was used in the implementation analysis. Implementation scores for each sponsor were determined by rank-ordering the non-Follow Through classrooms' mean scores on each sponsor variable and then dividing the distribution into five equal parts, or quintiles. Each sponsor classroom has a score on each variable and falls within a quintile range. A sponsor's implementation score on any variable is always a score between 1 and 5. This represents the position of a Follow Through classroom score relative to the distribution of non-Follow Through scores.

Using each variable designated as critical by the sponsor of a model, a total implementation score was computed for each classroom in each project location and for each sponsor. In order to assess the degree of implementation achieved by Follow Through classrooms, a total implementation score was also computed for each non-Follow Through classroom on each sponsor's set of implementation variables. The mean and standard deviation of the non-Follow Through pooled classrooms are reported for each sponsor separately for first and third grades. One-tailed "t" tests were computed to test for the significance of the differences between each Follow Through sponsor's classrooms and the non-Follow Through classrooms. Analyses of

Table 5-1

List of Critical Variables Selected by Sponsors

Variable	Far West Laboratory	University of Arizona	Bank Street	University of Oregon	University of Kansas	High/Scope	Education Development Center (EDC)
24, child selection of seating and work groups	X	X	X			X	X
25, games, toys, play equipment present	X	X	X			X	X
39, general equipment, materials present	X				X		X
65, guessing games, table games, puzzles		X			X	X	
66, numbers, mathematics, arithmetic	X	X	X	X	X	X	X[a]
67, reading, alphabet, language development	X	X	X	X	X	X	X[a]
70, sewing, cooking, pounding		X				X	
71, blocks, trucks						X	
74, practical-skills acquisition	X[b]		X				
83, wide variety of activities, over one day	X	X	X			X	X
86, teacher with one child	X	X	X			X	
87, teacher with two children			X				
88, teacher with small group		X	X	X	X	X	
92, aide with one child	X		X				
94, aide with small group		X		X	X	X	
114, one child independent	X	X	X				X
115, two children independent							X
116, small group of children independent	X		X				
239, mathematics or science equipment/academic activities	X		X			X	X
240, texts, workbooks/academic activities				X	X		
313, child to adult, all verbal except response			X				

344, individual child verbal interactions with adult	X	X	X	X	X	X	X
350, child questions to adults	X	X	X			X	X
363, child group response to adult academic commands, requests, or direct questions				X			
372, child presenting information to a group			X			X	
375, adult instructs an individual child	X						X
376, adult instructs a group				X			
390, adult task-related comments to children						X	X
394, all adult acknowledgment to children	X		X	X	X	X	
398, all adult praise to children		X		X	X		
412, adult feedback to child response to adult academic commands, requests, direct questions				X	X		
420, adults attentive to a small group	X			X		X	
421, adults attentive to individual children	X	X	X		X	X	
423, positive behavior, adults to children	X	X	X				
435, total academic verbal interactions							
438, adult communication or attention focus, one child	X		X		X	X	X
440, adult communication or attention focus, small group				X		X	
444, adult movement	X						X
450, all child open-ended questions							
451, adult academic commands, requests, and direct questions to children				X	X		
452, adult open-ended questions to children	X	X	X			X	X
453, adult response to child's question with a question						X	X
454, child's extended response to questions		X	X				X

continued

Table 5-1, *continued*

Variable	Far West Laboratory	University of Arizona	Bank Street	University of Oregon	University of Kansas	High/Scope	Education Development Center (EDC)
156. all child task-related comments	X	X				X	
157. all adult positive corrective feedback	X			X	X		X
160. all child positive affect	X	X					
469. all adult reinforcement with tokens					X		
509. child self-instruction, academic				X[a]			
510. child self-instruction, objects			X			X	X
513. child task persistence			X		X	X	
514. two children working together, using concrete objects						X	
515. small group working together, using concrete objects			X			X	
516. social interaction among children	X		X				X
574. child movement	X						X
598. child self-instruction, nonacademic	X	X				X	
Total *N* critical variables	28[b] 27[a]	21	27	16[b] 17[a]	17	29	20[b] 22[a]

[a] Third-grade only
[b] First-grade only

variance were also computed to examine the within-site and among-site differences in total implementation scores for each sponsor. Implementation is judged on two criteria: (1) Do the sponsored classrooms differ significantly from non-Follow Through classrooms? and (2) Are the classrooms similar in implementation both within projects and among projects? (See Stallings, 1975, p. 26, for this statistical procedure.)

RESULTS

The data obtained from this large sample indicated that the models in Head Start and Follow Through Planned Variation programs were very effective in training teachers in diverse locations to instruct in compliance with the models (i.e., Bank Street teachers in Tuskeegee looked similar to Bank Street teachers in New York City).

Further analyses of the observation data indicated that instructional processes identified with exploratory models explained 45 percent of the variance in scores on the Ravens Progressive Matrices. Instructional variables identified with direct instruction models explained 37 percent of the variance in reading achievement and 64 percent in mathematics achievement (Stallings, 1975). This was one of the first national evaluations of educational models to use a comprehensive observation system linking classroom processes to student outcomes.

OTHER STUDIES USING THE COI

Early Childhood Education

Following the initial study, the COI was used in a California study of an early childhood education program (Stallings, Cory, Fairweather, and Needels, 1979). The evaluation focused on the instructional processes of teachers in schools classified as having students with increasing achievement scores, compared with the instructional processes of teachers in schools where students' achievement scores were decreasing. This evaluation indicated more variance in instructional processes within schools than among schools. Overall

observation variables identified with direct instruction methods were significantly related to higher student achievement scores regardless of how the school had been classified.

Puerto Rican—English as a Second Language

The versatility of the coding system allows each coded interaction to be identified as English or Non-English. This capability was used in an evaluation of the quantity and quality of English being spoken in Puerto Rican classrooms. Puerto Rican observers were trained in a ten-day session to collect data reliably on the COI. Data were collected in urban and rural elementary and secondary schools, and observations occurred over two full days in randomly selected classrooms. The use of English was calculated as a percentage of the total recorded interactions. The quality of English used was assessed by reviewing tape recordings. (See Rivera-Medina, 1981.)

MODIFICATION FOR SECONDARY SCHOOL

For a study of teaching basic reading skills in secondary schools (Stallings, Fairweather, and Needels, 1978), the COI was modified to record the activities and instructional processes occurring in secondary classrooms. This study identified forty-one observation variables that were significantly related to a gain in reading achievement scores. Modifications in the coding and in the training program have been made to accommodate other subject areas such as science, social studies, mathematics, and physical education. For example, in a study of factors influencing women to take advanced mathematics classes (Stallings and Robertson, 1979), the COI was modified to identify when teachers were speaking to male or female students. The coding system provided variables that could be used to compare the nature of the interactions between male students and teachers with those between female students and teachers. Counter to our prediction, we found no significant differences in the classroom interactions among teachers and their male and female students. Because modifications to the program have occurred over the past ten years, we changed the name of the observation system to Stallings Observation

Table 5-2
Self-Analytic Model of Staff Development

Baseline/Pretest
- Observe teachers
- Prepare individual profiles of behavior
- Teachers assess what change is needed
- Teachers set goals
- Start where teachers are in skill development

Inform
- Provide information about research findings on effective practice
- Link theory, research, and practice
- Check for understanding by eliciting practical examples
- Ask, Why might that be? How does that work in your classroom?

Guided Practice: Integration
- Provide conceptual units one at a time
- Teachers adapt to own context and style
- Teachers assess and provide feedback via peer observations
- Teachers make a commitment to try a new idea in class the next day

Post-Test Observations
- Observe teachers: prepare second profile
- Teachers analyze profiles
- Teachers set new goals
- Assess training program for effectiveness

STAFF DEVELOPMENT BASED ON OBSERVATION

Phase I of the study of secondary basic reading skills was a year-long quasi-experiment in which very specific instructional variables were identified. Using these variables we constructed a staff-development program (named the Effective Use of Time [EUOT]). This training program, which was Phase II of the study and was based on an interactive theory of adult education, guided teachers to use the effective strategies. (See Table 5-2 for the model.)

In the year-long experiment, the teachers in the experimental classes successfully implemented the EUOT program, and their students gained six months more on reading achievement tests than did students in control classrooms. Findings from Phase I and II correlations and analysis of variance were remarkably similar. Summarizing the two data sets, we established the criteria shown in one teacher's profile (see Figure 5-2). These criteria then formed the basis for our recommendations for change in the teacher's behavior. The

Figure 5-2

The Percentage of Time Devoted by a Teacher to Certain Classroom Activities, in Relation to Established Criteria, with Recommendations for Change.

Activities	R*	Criterion***	Criterion Percentage	Teacher Baseline Percentage
Preparation				
Making assignments	Less	X	7	8
Organizing	Less	X	5	9
Teacher working alone	Less	X	3	15
Interactive Instruction				
Review/Discussing	More	X	10	6
Informing	More	X	20	14
Drill and practice	OK	X	2	2
Oral reading	More	X	9	2
*Noninteractive***				
Doing written work	OK	X	25	20
Silent reading	Less	X	9	20
Off Task				
Students socializing	Less	X	4	8
Students uninvolved	Less	X	5	15
Teacher disciplining	Less	X	1	6

* R= Recommendations

** Students work alone

*** Shows how much teachers exceeded (horizontal line to the right of vertical criterion line) or fell short of (line to left) the criterion for time devoted to the activity in question.

criteria are adjusted according to the achievement level of the students (Stallings, 1986).

DISSEMINATION

Federal and state education agencies, concerned for the many students in secondary schools who could not competently read, write, or compute, found the findings from the secondary reading studies of considerable interest. Subsequently, under the auspices of the Stallings Teaching and Learning Institute, the EUOT program was disseminated through the National Diffusion Network, and funding to assist in dissemination has continued from 1980 through 1990. The

EUOT program has been widely disseminated. Trainers for the EUOT program were certified at the Stallings Teaching and Learning Institutes at Vanderbilt University and the University of Houston. These participants included college of education faculty, school staff developers, and state department of education personnel. A body of research has evolved from these participants and from numerous student dissertations. The development of lap-top computer technology has allowed for more in-depth analysis and immediate feedback to teachers by providing instant profiles.

EUOT has been implemented in Branson, Missouri, over a three-year period. There, selected teachers were certified as trainers and observers to disseminate the program to all teachers in the district. Significant behavior change has been recorded for teachers and students throughout the project. In fall 1986, the Missouri Department of Elementary and Secondary Education identified Branson as a "Successful Project." Governor Ashcroft stated, "In the Branson school district, teachers and administrators have reported significant success with their Effective Use of Time (EUOT) program. This in-service experience helps teachers see how well they use class time and gives them strategies for using class time more effectively. The teachers involved in the program reported that EUOT helped them improve their skills significantly" (Orth, 1987, p. 4).

EUOT RESEARCH

Anderson (1984) examined the use of the SOS variables combined with Effective Use of Time (EUOT) training to improve instruction in the Washington, D.C., public schools. Her study focused on the changes in the teaching behavior of twenty-nine junior high school teachers who were trained by five different EUOT trainers. The SOS was used to determine the degree of change from the beginning to the end of the semester. The study examined the difference between the change of groups taught by four district trainers and one taught by an external trainer consultant. The teachers in the external consultant's group were found to change their behavior more than did teachers in the other groups. Anderson found that the most change occurred when the trainer (1) provided frequent teacher interactions, (2)

discussed the observation variables frequently, (3) made frequent supportive statements to try new ideas, and (4) stayed focused on the topic of the seminar (i.e., managing and motivation, student behavior, asking higher-level questions).

Longitudinal Study

Devlin-Scherer, Schaffer, and Stringfield (in press) conducted a follow-up observation study of an Effective Use of Time Program, for which they selected a sample of ten teachers who reflected high and low implementors from the original EUOT observations. They had three points of observation data (before the training, after the training, and two years later). High implementors had scores above the mean at the end of the training and the low implementors had post scores below the group mean. The ten teachers were observed and interviewed two years after receiving training in order to determine the long-term impact of the training on their teaching. The follow-up observations were compared with the initial observations on thirty-two variables. The average for the group of ten remained about the same over the elapsed time (i.e., change on eighteen of the observed variables was maintained at the same level as at the end of the training). On eight of the variables, the group's average was reverting to their initial behavior. Analyses of individual teacher's profiles revealed that teachers who had initially implemented the variables successfully were more likely to sustain their change than were teachers who implemented at a lower level. The high-implementing teachers indicated in interviews that the workshops provided them with present and future assistance. The low-implementing teachers indicated that sessions were confirmations of what they knew. They enjoyed the workshops more as opportunities to interact with peers. High implementors were able to identify specific skills they used in their classrooms. Low implementors were more global in their responses and less likely to identify specific skills.

Teacher Commitment

A study by Devlin-Scherer and colleagues (1985) entitled "The Effects of Developing Teacher Commitment to Behavioral Change" responds to the concern of measuring the effectiveness of training programs.

> Seventeen elementary and secondary teachers were trained by pairs of university, principal, or teacher trainers in the Stallings Effective Use of Time Program. Workshop sessions were audio-recorded and analyzed to determine the impact of verbal commitment behavior on changes in classroom teaching behavior. Using the SOS, a comparison of pre- and post-classroom observations indicated that teachers who stated public commitments to behavioral changes each week more often followed through and made these behavioral changes in their classroom teaching than did teachers who did not make such public commitment. [P. 31]

SOS FOR EVALUATION

Stringfield, Teddlie, and Suarez (1985) used the SOS to examine the classroom instructional processes of two Louisiana schools. One was identified as high achieving and the other was identified as low achieving. The majority of the students in both schools are white, and black, Hispanic, and Asian students form the minority population. Each school is located in middle-class, single-family-dwelling neighborhoods. The site team observers (who were blind to the achievement status of the schools) noted that students at the low-achieving school spent about one hour less a day doing academic tasks. During six days of observations, few classes began at 8:30 A.M. as scheduled. Many students were in the halls when the bell rang. The researchers indicated that there was "a constant stream of children to and from bathrooms, the office, the library, and in some cases, just hanging out in the halls" (p. 34). According to the SOS data, students in the high-achieving school received nearly twice as much interactive instruction as did students in the low-achieving school.

STUDENT TEACHING

A study by Harris (1988) included a sample of fifty student teachers. Over a fifteen-month period, twenty student teachers participated in full treatment of SOS feedback plus EUOT workshops and seven in feedback from SOS treatment only; there were twenty-three controls. Change was measured with eleven variables aggregating

subjects improved (moved toward criterion levels) for eight of eleven variables, with change significant at the 0.05 level for teachers' monitoring students, students in interactive instruction, and students off task. SOS feedback-only subjects improved for nine variables, with change significant for teachers' interactively instructing, teachers' managing, and students in interactive instruction. Implications for teacher education suggest that the feedback portion of the EUOT program is effective during the preservice teaching experience, that the portion of the EUOT workshops dealing with interactive instruction effects a change beyond that of SOS feedback only, and that trainer as observer increases student teachers' classroom management.

Freiberg and Waxman (1988) used three approaches for providing feedback to student teachers that have not been widely used but have great potential for improving the classroom instruction of preservice teachers. The methods include (a) feedback from pupils, (b) systematic feedback from classroom observation system (SOS), and (c) self-analysis of classroom lessons through an audiotape analysis (Low Inference Self-Assessment System; see Freiberg, 1987). The authors found that these feedback approaches, individually or collectively, provide student teachers, cooperating teachers, and university supervisors with excellent data for strengthening the preservice teaching experience. (See Freiberg, Waxman, Houston, 1987.)

Student Teaching in Inner-City Schools

The purpose of the Learning to Teach in Inner-City Schools project (LTICS) is to develop teachers who choose to teach in inner-city schools and are effective in teaching inner-city children. Historically, most new teachers did their student teaching in the suburbs. Those hired for inner-city schools had little preparation to serve children who come from a wide variety of cultural backgrounds and from low socioeconomic families. The dropout rate of new teachers in inner-city schools is reported to be twice the average for the nation (Stallings, Martin, and Bossung, in press).

The goal of LTICS is to change this history of failure by novice teachers in inner-city schools to one of success. To this end, a partnership was established between a school district serving inner-city students and a college of education that trains student teachers. The partnership created a professional development school that provides a structure in which a group of supervising teachers, college

supervisors, and ten to twelve student teachers per semester learn to implement effective instructional strategies for inner-city school populations. This occurs through shared required weekly seminars that follow the EUOT format. The seminars focus on the problems and solutions of teaching inner-city children (e.g., holding high expectations, working with parents and their children, assessing children's prior knowledge and experiences, planning appropriate lessons, managing classroom time, motivating and managing positive student behavior, and developing reflectivity and thinking skills). Seminars are taught by school and college faculty and community/parent representatives.

Student teachers and supervising teachers are observed with the SOS at the beginning of each semester and set goals for instructional change. The percentage of time children spend on academic tasks is computed and analyzed for change; these statistical analyses have indicated significant change each semester in student teacher and teacher behavior. The impact of LTICS is also evaluated by calculating the percentage of student teachers graduating from LTICS who choose to teach inner-city or other at-risk populations (85 percent at this time). Follow-up interviews with LTICS graduates indicate job satisfaction, and their principals give them high ratings.

SUMMARY

Observation in classrooms serves many purposes. Most often observation is used to evaluate teachers and students. The flexibility of the SOS has provided a means to identify effective instructional practices in a wide range of classroom settings. The specificity of the SOS variables and their face validity have made it relatively easy to translate them into teaching behaviors, and these data from the study have provided the content for extensive in-service and preservice professional development. The profiles of teaching behaviors observed in a pretest and posttest design provide a continuing basis for evaluation and improvement of the EUOT program.

REFERENCES

Anderson, Sandra Lee. "Teacher Training Techniques from Four Observational Perspectives," *Journal of Classroom Interaction* 20, no. 1 (1984): 16–28.

Deutsch, Martin. *The Disadvantaged Child*. New York: Basic Books, 1967.

Devlin-Scherer, Roberta; Devlin-Scherer, Wade; Schaffer, Eugene; and Stringfield, Samuel. "The Effects of Developing Teacher Commitment to Behavioral Change," *Journal of Classroom Interaction* 21, no. 1 (1985): 31–37.

Devlin-Scherer, Roberta; Schaffer, Eugene C.; and Stringfield, Samuel C. "A Two-Year Follow-up of a Staff Development Program Designed to Change Teacher Behavior," *Journal of Classroom Interaction*, in press.

Egbert, Robert L. "Planned Variation in Follow Through." Paper presented at the Brookings Institute Conference on Social Experimentation, Washington, D.C., April 1973.

Flanders, Ned A. *Analyzing Teacher Behaviors*. Reading, Mass.: Addison-Wesley, 1970.

Freiberg, H. Jerome. "Teacher Self-evaluation and Principal Supervision," *National Association of Secondary School Principals Bulletin* 71 (1987): 85–92.

Freiberg, H. Jerome, and Waxman, Hersholt C. "Alternative Feedback Approaches for Improving Student Teachers' Classroom Instruction," *Journal of Teacher Education* 39, no. 4 (1988 : 8–14.

Freiberg, H. Jerome; Waxman, Hersholt C.; and Houston, W. Robert. "Enriching Feedback to Student Teachers through Small Group Discussion," *Teacher Education Quarterly* 14, no. 3 (1987): 71–82.

Harris, Alene Hawes. "Sources of Treatment Effects in a Teacher Effectiveness Training Program." Doctoral dissertation, Vanderbilt University, August 1988.

Orth, Lee J. "A Good Time Was Had by All." In *School and Community*. Branson: Missouri State Teachers Association, 1987.

Rivera-Medina, Eduardo. *Assessment of the In-service Training Needs of Teachers of English and of Spanish to Returned Migrants in Puerto Rico*. Final Report. Washington, D.C.: National Institute of Education. 1981.

Stallings, Jane A. *Follow Through Program Classroom Observation Evaluation, 1971–1972*. SRI Project URU-7370. Menlo Park, Calif.: Stanford Research Institute, August 1973.

Stallings, Jane A. "Implementation and Child Effects of Teaching Practices in Follow Through Classrooms," *Monographs of the Society for Research in Child Development* 40 (1975), Serial No. 163.

Stallings, Jane A. "Effective Use of Time in Secondary Reading Programs." In *Effective Teaching of Reading: Research and Practice*. Newark, Del.: International Reading Association, 1986.

Stallings, Jane A.; Baker, Phil; and Steinmetz, Gerald. "What Happens in the Follow Through Program: Implications for Child Growth and Development." Paper presented at the meeting of the American Psychological Association, Honolulu. September 1972.

Stallings, Jane A.; Cory, R.; Fairweather, James S.; and Needels, Margaret. *Early Childhood Education Classroom Evaluation*. Final Report for the Department of Education, State of California. Menlo Park, Calif.: SRI International, 1979.

Stallings, Jane A.; Fairweather, James S.; and Needels, Margaret. *A Study of Teaching Basic Reading Skills in Secondary Schools*. Final Report for the National Institute of Education. Menlo Park, Calif.: SRI International, 1978.

Stallings, Jane A., and Gieson, Philip. "The Study of Reliability in Observational Data." In *Phi Delta Kappa Occasional Paper*, no. 19. Bloomington, Ind.: Phi Delta Kappa, 1977.

Stallings, Jane A.;Martin, Anna; and Bossung, Joan. "Houston Teaching Academy: A Partnership in Developing Teachers," *Teaching and Teacher Education*, in press.

Stallings, Jane A., and Robertson, A. *Factors Influencing Women's Decisions to Enroll in Advanced Mathematics Courses*. Final Report for the National Institute of Education. Menlo Park, Calif.: SRI International, 1979.

Stringfield, Samuel C.; Teddlie, Charles; and Suarez, Sandra. "Classroom Interaction in Effective and Ineffective Schools: Preliminary Results from Phase III of the Louisiana School Effectiveness Study," *Journal of Classroom Interaction* 20, no. 1 (1985): 31–37.

Wolff, M., and Stein, A. "Study I: Six Months Later—A Comparison of Children Who Had Head Start Summer 1965 with Their Classmates in Kindergarten." (A Case Study of Kindergartens in Four Public Elementary Schools, New York City). Yeshiva University, August 1966.

TEMPLE UNIVERSITY CENTER FOR RESEARCH IN HUMAN DEVELOPMENT AND EDUCATION

The Temple University Center for Research in Human Development and Education is an interdisciplinary center for the study of emerging problems and challenges facing children, youth, and families. Its overall goal is to investigate the basic forces that affect human development as well as educational processes and outcomes. An important focus of the Center's work is the identification and shaping of effective responses to these forces through far-reaching changes in institutional policies and practices.

The problems and challenges facing children, youth, and families stem from a variety of cultural, economic, political, and health pressures. Their solutions are, by nature, complex. They require long-term programs of study that apply knowledge and expertise from many disciplines and professions. To this end, the Center draws together the many resources of the University and a wide range of national, state, and regional programs. The result is interdisciplinary and interdepartmental collaborations that involve investigations of social, economic, educational, and developmental factors and demonstration of state-of-the-art models for training and for delivery of relevant services. Research and development projects in these areas reflect a commitment to enhance the knowledge base for improving the quality of life for children and families, particularly in urban environments.

The work of the Center for Research in Human Development and Education is divided into four program units: Improving Instruction and Learning in Schools, which provides technical assistance and training for innovative school programs; Social Service Delivery Systems, which develops models for effective social service delivery; Studies of Child Development and Early Intervention, which conducts pre-school diagnosis and produces innovative program development; and the National Center on Education in the Inner Cities (CEIC), funded by the U.S. Department of Education's Office of Educational Research and Improvement, which has undertaken a program of research and development as well as dissemination that takes bold steps to mobilize and strengthen education and related resources to foster resilience and learning success of children, youth, and their families in inner cities.

The Center is supported by funds from Temple University as well as by grants from public agencies and private foundations. The following are the current funding sources:

Bell Atlantic
Ben Franklin Partnership Program of the Advanced Technology Center for S. E. Pennsylvania
Carnegie Corporation
City of Philadelphia
- Department of Human Services
- Mayor's Commission on Literacy
- Office of Mental Health and Mental Retardation
- School District of Philadelphia

Commonwealth of Pennsylvania
- Department of Education
  - Bureau of Special Education
  - Bureau of Vocational and Adult Education
- Department of Labor and Industry
- Department of Public Welfare

Delaware State Department of Education
Exxon Education Foundation
Geraldine R. Dodge Foundation
IBM Corporation
Louisiana State Department of Education
National Science Foundation
Private Industry Council of Phila.
Rockefeller Foundation
The Pew Charitable Trusts
United States Department of Education
- Office of Educational Research & Improvement
- Office of Elementary & Secondary Education
- Office of Special Education Programs

United States Dept. of Health & Human Services
United States Department of Labor
William Penn Foundation
William T. Grant Foundation

For further information, please contact:

Temple University Center for Research
in Human Development and Education
13th & Cecil B. Moore Avenues
Philadelphia, PA 19122
(215) 787-3000